How to translate artificial intelligence? Myths and justifications in public discourse

Jonathan Roberge, Marius Senneville and Kevin Morin

Abstract

Automated technologies populating today's online world rely on social expectations about how "smart" they appear to be. Algorithmic processing, as well as bias and missteps in the course of their development, all come to shape a cultural realm that in turn determines what they come to be about. It is our contention that a robust analytical frame could be derived from culturally driven Science and Technology Studies while focusing on Callon's concept of translation. Excitement and apprehensions must find a specific language to move past a state of latency. Translations are thus contextual and highly performative, transforming justifications into legitimate claims, translators into discursive entrepreneurs, and power relations into new forms of governance and governmentality. In this piece, we discuss three cases in which artificial intelligence was deciphered to the public: (i) the Montreal Declaration for a Responsible Development of Artificial Intelligence, held as a prime example of how stakeholders manage to establish the terms of the debate on ethical artificial intelligence while avoiding substantive commitment; (ii) Mark Zuckerberg's 2018 congressional hearing, where he construed machine learning as the solution to the many problems the platform might encounter; and (iii) the normative renegotiations surrounding the gradual introduction of "killer robots" in military engagements. Of interest are not only the rational arguments put forward, but also the rhetorical maneuvers deployed. Through the examination of the ramifications of these translations, we intend to show how they are constructed in face of and in relation to forms of criticisms, thus revealing the highly cybernetic deployment of artificial intelligence technologies.

Keywords

Translation, justification, criticisms, cybernetics, governmentality, artificial intelligence

> This article is a part of special theme on The Turn to AI. To see a full list of all articles in this special theme, please click here: https://journals.sagepub.com/page/bds/collections/theturntoai

Introduction

It has already been a few years since media and pundits of all sorts announced the onset of a full-fledged "AI revolution," the result of a noticeable effervescence of newly developed machine learning techniques as well as the urgency of many firms to weave their ever-smarter innovations into the fabric of everyday life. And indeed, in this regard the field of artificial intelligence (AI) has made phenomenal strides, not only in the digital realm of platforms but in our material, immediate environments, which have become increasingly subject to the same bleakly inductive and classificatory matrices. Conversely, the social sciences and humanities have been characteristically late to the party. Slow to react to these ongoing transformations, they could, for the most part, still be portrayed as disorganized and unprepared for what awaits. The question then is what would be required to develop better, more encompassing and programmatic views? A first set of such conditions deals with the contextualization

Institut national de la recherche scientifique, Quebec City, Canada

Corresponding author:
Jonathan Roberge, 490 Couronne St., Quebec, QC G1K 9A9, Canada.
Email: jonathan.roberge@ucs.inrs.ca

of both our disciplines and the technologies they are meant to assess. Fields such as Internet, Network, and New Media Studies have successively transitioned their focus from online sociability and Big Data to algorithms and algorithmic cultures (Roberge et al., 2019; Roberge and Seyfert, 2018). Today's artificial neural networks and other machine learning variations all borrow from the above objects of study; yet they do so unevenly and with the unnerving habit of creating flexible—if not *ad hoc*—assemblages that are difficult to pin down. Backpropagation, recursive loops, and other such subtleties are shaping new data architectures from the inside-out while also interacting and coping with data originating from the outside—social—world (L'Heureux et al., 2017), thus once again proving how critical it is to never lose sight of the contexts in which those systems operate.

Closely related to the idea of contextualization are the issues of interdisciplinarity and theoretical cross-pollination. Correctly accounting for these requires that we acknowledge how, up to now, significant fields such as Science and Technology Studies (STS) have developed in contrast to, rather than in dialogue with, former paradigms. Bruno Latour's work, to focus on a preeminent example, has for the most part been construed in opposition to the hermeneutical or critical orientations offered by Paul Ricœur or Jürgen Habermas, for instance.[1] The issue with such quarrels, of course, is that they do no service to actual research. There is an urgent need today to contextualize, i.e., historicize and culturally situate the conduct of STS. Authors and concepts that are seldom assembled can and should be in order to better understand and decipher the influence of AI on society and vice versa.[2]

The concept of *translation* will herein be a first case in point. Developed by Callon and, to a lesser extent, Latour and others (Akrich et al., 2002a, 2002b; Callon, 1984; Latour, 2005), it has traditionally been used in regards to what happens when an object of science moves from one state or situation to another—a relevant example being formal mathematics converging into today's advances in deep learning (Mackenzie, 2017). From a beta project in computer science to the mass distribution of a commercialized product, not only has the innovation been micro-adjusted to the requirements of the outside world, but the outside world itself has been fine-tuned to the incoming novelty. When AI disseminates throughout all parts of society, it is through such translations that explanations about what it does (or could soon be doing) are built up as mainstream knowledge. First section will discuss this in more detail, but suffice to say for now that the notion of translation could account for the whole range of hype, myth, prediction, and symbolic ambiguity—as in more hermeneutical and critical traditions (Roberge,

2011, 2016). Yet translations should never be accounted for as disembodied processes playing out in the ether; such displacements of meaning cannot be accurately rendered without proper attention given to the translators. Who speaks? Through which channels and with how much success? How well, to refer back to a central notion of the sociology of translation, does one manage the art of *problematization*, where one intends to predefine the issues to be resolved and thus propose a certain understanding of the roles of all the actors involved (Callon, 1984)? Throughout this contribution, we want to argue (i) that such a *translation-translator-problematization tryptic* allows for a new understanding of the ways through which AI is brought to the public sphere, showing how fundamentally unequal translators are and how the process of problematization presently deployed is in fact its own—very much partial and regional—solution. For translators, supporters, and the like, the act of raising questions about the limits and acceptability of AI constitutes the answer in what is thus a contextualized, performative, and yet understudied endeavor.

AI, as an object of historical inquiry, has always existed in a state of public controversy. Thus, it becomes all the more important to pay attention to the discourses and rhetorics displayed by both its proponents and opponents. What are the tropes, metaphors, and analogies being used—and again, to what outcomes? By raising such questions, it becomes possible to bridge STS with other, more recent theoretical currents that focus on legitimation processes (Boltanski, 1990; Lash, 2007). Translation in many ways coincides with such a concept, in that they both emerge and coalesce *in face of* and *in relation to* forms of criticism that themselves are contextual and performative. Mobilizing recent examples of various kinds and stakes, we want to argue (ii) that the *current unfolding of machine learning systems benefits from a diffuse, if not latent, critique* compared to its increasingly dense and fused justification. At this point, the first section will resume by further developing the underlying dynamic of today's mass deployment of AI technologies, namely, what it incurs in terms of rationalities and logics of control. Such a deployment should be understood as generative or performative, in the sense that it revolves around ambiguous, yet all the more efficient and self-fulfilling, predictions. Concurrently, this represents the third major argument we want to put forward, namely, (iii) that what is to be witnessed today is *the rise of a highly cybernetic form of power* by, through, and in favor of a set of technological developments hazily called "AI."

This contribution is organized as a rather substantial first theoretical section, followed by three smaller and more empirical sections, each presenting specific

case studies that are emblematic of current developments in the research, economic, and political realms. We thus intend to echo the call made by prominent scholars regarding the importance of undertaking more *in situ* investigations (Kitchin, 2014; Mackenzie, 2017).Combining this empirical focus with Michel de Certeau's notion of *un multiple* he developed in *Culture in the Plural* (1974), our goal is to emphasize how each of these three cases illustrates the way AI emerges both ontologically and epistemologically as its own archipelago: each of its deployments can be shown to be all at once multiple, disparate, and yet commensurable and interrelated. In second section, we analyze the Montreal Declaration for a Responsible Development of Artificial Intelligence as a prime example of how stakeholders manage to establish the terms of the debate by preaching the gospel of ethical AI while simultaneously avoiding substantive commitment. Third section focuses on Mark Zuckerberg's 2018 congressional hearings, inasmuch as they offer a crystallization of the ways the company is capitalizing today on machine learning as the solution to its many problems. Zuckerberg, hardly the most charismatic orator, nonetheless managed to give sufficiently vague and, *in fine*, performative answers so as to circle around the poorly framed criticisms of his interrogators. Finally, the fourth section focuses on both the way actors in AI engaged in various problematizing endeavors aimed at redefining the normative landscape around autonomous weapon systems and the different elements of explication accounting for the failure of the 2018's Campaign to Stop Killer Robots.

Theoretical architecture: Translation, justification, and cybernetics

It might not be possible—nor helpful—to describe all the details and ramifications of Callon's and Latour's sociology of translation. At its simplest, translation refers to the act of bridging and linking, of "creating convergences and homologies by relating things that were previously different" (Callon, 1981: 211).In the realm of discourse, translation partakes in meaning-making through the dissemination of common and readily usable languages. In the realm of action, translations allow for the connection and bidirectional influence of actors—be they nonhuman "actant"; human-like, such as a large organization; or "truly" human, as in the case of a leader or a spokesperson embodying a broader assemblage. Here lies the "network" quality of Callon and associates' STS approach, in that it can account for both the quasi-tectonic displacements of underlying currents as well as the more personalized dynamics that occur. "The fate of innovation," they

say, "rests entirely on the choice of the representatives or spokespersons who will interact, negotiate to give shape to the project" (Akrich et al., 2002b: 217).

The so-called "Godfathers of AI" and latest recipients of the 2018 Turing Award—namely, Geoffrey Hinton, Yann LeCun, and Yoshua Bengio—could be taken as prime examples of what is meant by the notion of translator (Metz, 2019). The idea that they "work together" is especially relevant as they symbolically rely on each other to disseminate their core values and views about connectionist machines, prediction through induction, and so on (see Cardon et al., 2018). Yet, this common endeavor still translates into drastically different styles of attending to the task of problematization, each corresponding to the various degrees with which said researchers intend to engage society on the various issues raised by AI implementation. Of the three, Bengio is, without a doubt, the one most invested in social and political matters. He was instrumental in the mobilization of both political and civil society actors around the Montreal Declaration, and he is an important figure in the movement against so-called "killer robots," two topics we will address shortly. For now, let us simply emphasize how both of these ventures resulted in significant gains in notoriety as well as a remarkable escalation of Bengio's role as both broker and problematizing figure in the field of AI.

"To adopt an innovation is to adapt it" should be considered a mantra in the sociology of translation (Akrich et al., 2002b: 209).Central, then, is this process of *problematization*, whereby questions of who is involved, what is going on, when and why it occurs, etc., should at least be tacitly addressed. Translators are those who achieve certain success in identifying relevant actors, mobilizing what they are about, and identifying what they can contribute—thus partaking in a broader legitimization effort. Is such effort strategic? Is it highly contextualized? It is to be sure; as Akrich et al. observed, "as in all negotiations, the possible strategies are infinitely varied. The means to be mobilised in order to arrive at a compromise are a *matter of circumstance*" (2002b: 210; emphasis added). The current AI deployment represents a prime example of this dynamic. On the one hand, it demonstrates how today still only a handful of translators have any practical influence: renowned scientists but also larger corporations such as Google or Facebook, often with the support of state institutions investing financial and/or symbolic resources aimed at bolstering AI's social acceptability.[3] On the other hand, special attention must be paid to the "regional" circumstances that make any case unique. To give the example of Quebec, the province relies on a four-pillar scheme: an industrial cluster subsidized by the government, the *Cité de l'IA* in Montreal,

an International Observatory on the Societal Impacts of AI, and an ethical Declaration—all of these in one way or another associated with Bengio (Roberge et al., 2019). Moreover, both strategic and contextual elements have the ability to show how, in the case of AI deployment, problematization is itself a particular brand of "solutionism" (Morozov, 2013). The pitfalls, risks, and dangers are barely stated, and when they are, it is mostly to signify a vague sense of awareness.

Translations are as performative as they are stated for an audience to hear and make sense of. Broader echoes and decipherments are of the utmost importance in that they could allow for what is perhaps a more "culturalist" follow-up to Callon's theory. "[I]ndeed, it would be hard to picture the formation of technology developments and innovation without some kind of shared, though flexibly interpreted, cluster of guiding visions," Borup et al. (2006: 289) note. Speculations, desires, expectations, and the like surge in all directions, motivated by powerful and often emotionally charged rhetoric. As a whole, these come to inform broader imaginaries and narratives that are truly mythical in nature (Elish and boyd, 2018). AI currently enjoys a profound as well as multifaceted hype that might be rooted in the sort of ambiguity that comes with an uncertain and contingent future.[4] Hype, ambiguity, and efficiency go hand in hand. "Hype is low on informative content," Guice writes, "but directly states the relevance of the information to a social context" (1999: 85).

As stated above, debates are a constitutive part of complex phenomena such as the deployment of AI. Success is not defined as the absence of criticism but as its adequate handling. Following Natale and Ballatore, "skepticism and criticism [have] added to AI's capacity of attracting attention and space [...] in the public arena" (2017: 2). What there is, in other words, is a subtle yet forceful dance between the capacity to denunciate and the capacity to justify the technology. From a theoretical point of view, this could be related to Boltanski's sociology of critical capacities and how, for some, it represents the "symmetrical twin" of Latour and Callon's theoretical frame (Boltanski, 1990; Boltanski and Thévenot, 2006; Guggenheim and Potthast, 2012).Whereas criticisms come forward in the public discourse questioning any potential flaws, justifications are offered as discourses aimed at appealing to higher principles their proponents hope will command respect—innovation, progress, etc. For Boltanski, criticism and justification are inseparable; as mentioned, they are constructed *in face of* and *in relation to* one another. Yet, this is not the same as making the argument that they are on par. This is certainly crucial in order to understand the current state of affairs in the deployment of AI. Who are

the translators of the critique? Where do they operate? And which practical and symbolic modes of influence are they using? By comparing the critical voices speaking against the deployment of AI technologies to those of star researchers like LeCun or Bengio, powerful corporations such as Facebook or Amazon, but also state actors, one quickly concludes that they simply don't compare. Yet again, this is not to say that criticism has vanished. What it means is that it exists either in a form of latency—the distant fear of the Singularity or Skynet, for instance—or in a scattered, patchy mode in which arguments and/or actors have great difficulty connecting and coordinating. The different cases we discuss in the second, third and fourth sections illustrate this contextual yet encompassing logic, whereby criticism is more or less neutralized, if not recycled, by justificatory discourses. As it will appear more clearly, in this decisive *pas de deux*, justification ends up leaving criticism very little leeway.

Legitimacy through performance, problematizations becoming practical solutions, and criticisms co-opted by justifications all come to inform a brave new world powered by AI that is *ad hoc*, in flux, and adaptive. In a nutshell: such deployment has come full circle in a way that is highly reminiscent of Robert Merton's (1948) idea of "self-fulfilling prophecies." This paper builds on this key notion while adding to it through theoretical refinements stemming from cybernetics—i.e., the broad idea that control operates through communication and feedback in order to produce changes in the environment. Halpern, for instance, describes the sort of rationality set in motion in systems that "can both recognize and disavow their history [and which are] incarnated in a drive to accelerate the speed of speculation while intensifying the infrastructure for data gathering and storage" (2014: 236). Artificial neural networks provide a telling illustration of this logic. Their entire architecture and *modus operandi* rely on backpropagation mechanisms, namely, the ability to account for errors in the initial output of the model as signals that allow for a constant readjustment of the parameters employed in the ensuing processes (Mackenzie, 2017). Such powerful tools have induced the emergence of a substantial justificatory and normalizing assemblage, rendered as a whole cottage industry of blogs, briefs, and promotional kits: "AI software can take what looks like a bunch of unmanageable, noisy data," writes Gagné (2017), "and turn it into something useful: signal. Noise is becoming manageable information, and there is a lot of noise to play with". Such a swirling and looping-back dynamic thus informs every step and aspect of AI progression. As a distinctive technoscience, an industry, and a field of actors and meanings, AI and its reign are best characterized as fully cybernetic.

Heavy emphasis is placed on lightning-fast speed and nimble iterative production processes that rely on data-driven decision making, as both scalability and open-endedness are at a premium in a world progressively dominated by A/B testing governance. In ever-wider contexts, a similar ethos is beginning to hold sway, increasingly determining the ways to conduct business, as well as be in and approach the world.

To what extent is it then possible to describe such a new, broad, and cybernetic mode of social engineering? The issue is of the utmost importance as discursivities, rationalities, and logics of power and control are all at stake. The end results of such developments are unavoidably political and normative—facial recognition systems being a prime example. Big Tech firms should primarily be understood as optimizing, all-encompassing agents: the social problems and issues they induce by massively implementing AI technologies are envisaged as outputs, noise, and potential signals that might be used to readjust their whole technological display. Focusing at the outset on scalability and emergent effects, they deploy first and monitor, justify, and fine-tune later. As will be shown in the third section, most of their interactions with society at large take the form of trial balloons, i.e., fuzzily delineated experiments intended to gather information on whatever grounds they can cover without encountering too much resistance. Certainly, in a world where success and advancement are based on flexibility and mobility—and where prediction, control, and ambiguity go hand in hand—power lies in informed adaptation.

Translators and the self-regulating horizon of ethical AI

There is now a cottage industry of declarations, frameworks, roadmaps, boards, and the like dedicated to the handling of ethical matters in AI: cities like Montreal and Toronto have given their names to the "Declaration for a Responsible Development of Artificial Intelligence," and the "Declaration Protecting the rights to equality and non-discrimination in machine learning systems"; Open AI, sponsored by Elon Musk, currently works toward safe Artificial General Intelligence; and Amazon, Apple, Google, Facebook, IBM, and others are promoting the "Partnership on AI to Benefit People and Society." The question then is how to make sense of these efforts, namely, what is it that they have in common that enables them to occupy such a central stage in this day and age? Following the theoretical considerations laid down in the first section, it certainly is possible to argue that what we are witnessing are translations turning into problematizations and then developing into justifications. Fears are to be addressed, interpretations are to be tempered, and sensitivities are to find a language by which they can be communicated. As Greene et al. note, "building a moral background for ethical design is partly about shaping public perception" (2019: 8). And yet, it is precisely the conceptual and enunciative nature of the language used that also marks the limitation of these ethical discourses—their very structure as "speech acts" being both what shows *and* hides (Ricœur, 1998).A distinctive feature of today's ethical AI cottage industry is indeed the vagueness of the terms proposed: well-being, transparency, and so on. Ambiguities help the flow of ideas and, in that sense, are both performative and adaptive. Moreover, it is such vagueness that allows for engagement in specific contexts. As will become apparent below, local ethical AI initiatives are of interest because they most often point toward global developments they have little control over. The way they deal with their given constraints—legitimating public investment in a local ecosystem for instance—is to operate a sort of *montée en généralité* that in itself is an attempt to avoid criticism.

A key example of such a "glocalized" and all-encompassing ethical statement is the Montreal Declaration, drafted by a group of researchers mostly from the Université de Montréal and, in reality, the concretization of Bengio's moral vision. While context matters, it might be important to recall that Quebec is a rather small and distinct society where the state occupies a central role. The Declaration, as said, is part of a larger assemblage that also comprises a state-supported industrial cluster, a district in Montreal where public and private investment is focused, and a nascent Observatory on the societal impact of AI. Institutionalization and personalization not being in contradiction, all these initiatives are Bengio-related and all have benefited—and still benefit—from unflagging state support (Roberge et al., 2019). The Declaration itself has been time and again promoted by Quebec's Chief Scientist's Office, while much at the same time the province's Minister of Economic Development has made clear the strategic importance of the industry: "Artificial intelligence certainly is a priority. I think too often we've sprinkled [subsidies] in Quebec and that we do not clearly choose what is important for us. We're now making a very concrete commitment" (Rettino-Parazelli, 2017; our translation).Thus, we see how scientific, economic, and ethical justifications were articulated to ensure the public's embrace of this booming new industry.

The Declaration is the result of a two-year process of consultation with diverse actors from public and private sectors. From the outset, three objectives guided its elaboration: "to create an ethical framework

for the *development* and *deployment* of AI"; "to guide the *digital transition* so that everyone can benefit from this *technological revolution*"; and "to open a space for *national* and *international dialogue*" (IA Responsible, 2018; our emphasis). Herein, the *development* of AI technologies is presented as inevitable, much like the *technological revolution* is treated as a teleological certainty. Problematically, those guiding principles don't mention or plan to question whether those technologies are safe, should be developed, or whether certain surveillance technologies should be made illegal. Furthermore, the *national* and *international dialogue* they refer to point to the way the current deployment of AI technologies is not a matter that can be solely handled in local political contexts, paradoxically revealing the very limits of their own endeavor.

Digging deeper in the Declaration—and what it obfuscates—epistemological issues begin to appear. The approach is one said to be of "co-creation" and "co-construction," an increasingly popular notion nowadays despite the fact that it has yet to find a solid rationale in the literature. To the contrary, such a method allows the work and expertise that social scientists can provide to be short-circuited, when they are not simply admonished and dismissed as suffering from "ivory tower" syndrome. These discussions can still be fruitful, but they run the risk of being less informed by input from historical, legal, social, or economic analyses. Indeed, the circumvention of social science expertise—i.e., those scholars willing to adopt an agonistic posture (Crawford, 2016)—means that a robust checks and balances mechanism is bypassed. For its part, the Montreal Declaration, by design, escapes such a potential threat. While it prefers to build on what *ought* to be, little energy is dedicated to what is currently known about AI deployment. Legitimacy is thus outsourced to an abstract collective as a soft consensus appears to be the solution coming from Quebec.

Problems continue to proliferate if one turns from how to achieve consensus to what the actual consensus says. The fact of the matter is that the Montreal Declaration's value statements and core principles are so broad that they do not even specifically address AI. A sample of such "values" could include for instance: "The development and use of [AI] should increase the welfare of all sentient beings"; "The development and the use [of AI] must contribute to the realization of a just and fair society"; or Principle 6.2—probably the most self-contradictory—"AI development must help eliminate relationships of domination between groups and people based on differences of power, wealth, or knowledge" (IA Responsible, 2018). The fact that the automation of knowledge production is the sole definition of machine learning, the reason why people invest time and money in the technology in the first place, and the way that firms are able to create such metamorphoses within power and capital distribution is simply ignored. In turn, such a "performative contradiction" creates a semantic network of terms that more often than not are self-referential and free-floating.

As was emphasized earlier, the broad, international, and "meta" nature of the Montreal Declaration also points out that it is not a matter that can realistically be handled in a local political context. The context, in other words, is one that makes the context apparently not very relevant—the (American) Partnership for AI struggles somehow conversely to acknowledge how American it actually is in its views and objectives. Looking at the Quebec government's efforts, one sees that it has done nothing but promote AI and invest in Montreal's AI ecosystem. The mantra is *governance at a distance*: to navigate the complexities of the present, it is deemed better to aim for a horizon that is as remote as possible and hope for the best. The Quebec government has, for example, no data privacy reform planned, nothing to say about facial recognition technologies, and only a vague understanding of the consequences AI will have on the job market and higher education. Its wait-and-see approach thus remains for the most part silent on practical issues and largely dependent on what happens politically and economically in Europe and the United States (with regards to the evolution of the EU's General Data Protection Regulation (GDPR), for instance). In this purportedly decontextualized and highly cybernetic context, the Montreal Declaration appears to fall between what Wagner (2018) calls "ethics-washing" and "ethics-shopping.". Selective and charitable principles inform a patchwork that somehow shields itself from any checks and balances, regulations, constraints, or opposition. But are these not crucial in preventing bias, prejudice, and moral scandals in AI? So far, as seen in the Montreal Declaration as well as Google's latest debacle with its short-lived External Advisory Council for AI Ethics—which included the CEO of a drone company and the president of a right-wing thinktank—such self-administered medicine appears to be far more hazardous (Levin, 2019).

Machine-Learned solutionism and justificatory performance: Zuckerberg's testimony

In early April 2018, Facebook CEO Mark Zuckerberg appeared in public hearings held by the US Congress in what was shaping up to be one of the defining moments in the company's history. At the center of the event

were allegations regarding Cambridge Analytica's utilization of over 87 million Facebook users' personal data during Trump's 2016 presidential run to create so-called "psychographic" profiles of individual voters (Rosenberg et al., 2018). Most media reported on the tragicomic aspect of the hearings, as Zuckerberg's uneasiness and definitive lack of bravado were counterbalanced by the quasi-comical technical ignorance revealed by some of the elected officials (Newton, 2018). Yet, a more critical and robust investigation would have noted the way responsibility for the whole debacle was twice deferred. First, elected officials appeared far more inclined to listen to Zuckerberg's opinions on regulatory matters than to actually undertake a leading role in the talks. And second, the not-too-charismatic CEO himself repeatedly dodged questions regarding accountability and instead promoted machine-learned content moderation systems as *the* solution to the whole debacle. The technical limitations of AI-gatekeeping were played down by the sure-to-happen technological advancements already in the pipeline. All in all, Zuckerberg's testimony is thus highly emblematic of a symbolically charged, technocratic, and cybernetic "ritual of power"[5] that blends prediction and prophecy, thus allowing its main protagonist—i.e., translator—to gain agency, i.e., time to enact the promised salvific changes and political room to maneuver accordingly.

Throughout his two-day-long testimony, Zuckerberg (2018b) put much emphasis on what he clearly envisioned as his favored solution, namely, an ever-increasing use of machine-learning systems deployed to identify and, as much as possible, preemptively root out harmful content: "the sheer volume of content on Facebook makes it so that [no amount] of people that we can hire will be enough to review all of the content. We need to rely on and build sophisticated A.I. tools that can help us flag certain content." If such a massive implementation of automated decision-making systems—admittedly helped, in the first phase, by more than 20,000 human moderators—appears to be the sole scalable answer, many experts argue that both text and image processing still lack the capacity to correctly decipher contexts and meanings (Gillespie, 2017). These technical objections, acknowledged by Zuckerberg (2018b) during his testimony—"Some problems lend themselves more easily to AI solutions than others. So hate speech is one of the hardest, because determining if something is hate speech is very linguistically nuanced, right?"—were nonetheless circumvented through the sheer promissory value of what Elish and boyd have called "the magic of Big Data and AI": "By calling upon a future that is imminent but always just beyond reach, what technologies can currently do is not as important as what they

might do in the future" (2018: 10). One couldn't better describe Zuckerberg's (2018a) rhetoric as, for example, when he contended: "I am optimistic that, over a 5 to 10-year period, we will have A.I. tools that can get into the linguistic nuances of different types of content." Through repetitive and intentionally vague references, and without ever touching on technical specificities, the Facebook CEO thus managed to evade any interrogation or attack directed at him.

But what would it mean for Facebook, YouTube, or any other digital platform to actually develop and implement effective machine learning systems of content moderation? AI-gatekeeping would have to be interpreted as quite the strategic adjustment when considered in terms of what Gillespie (2010) has called platforms' "sweet spot." Digital platforms, he explains, have historically attempted to preserve a neutral "honest broker" and "hands-off" posture—relying on the safe harbor legislation that protects legally defined "online intermediaries." Faced with the magnitude of the events for which digital platforms are now being held responsible—a telling example being the recent use of Facebook by the Myanmar military in their operations against Rohingya populations—such a position might simply not be possible to hold anymore. Departing from their justificatory traditional repertoire, Facebook's representatives now have to achieve a new balancing act between the liability-free stance they are leaving and the full-on responsibility they are still actively trying to avoid. This repositioning has been carried out through a two-pronged legitimacy claim: to conduct a "large-scale democratic process" aimed at the identification of culture- and country-specific moderation guidelines—"The idea, says Zuckerberg, is to give everyone in the community options for how they would like to set the content policy for themselves"—which are then to be enforced through machine-learned tools (Zuckerberg, 2017). The democratically informed, yet automated and therefore neutral, representational quality of such a process is what affords Facebook the necessary margin for its new, if vague and somehow depoliticized rhetorical posture. AI-gatekeeping, portrayed as the efficient and scalable solution to and impartial arbitrator of what is or isn't harmful content, thus allows for the depoliticization of the editorializing enterprise and reconciles Facebook with the "online intermediary" status it is aiming for—and the relative partisan immunity that extends from it.

The noticeable emphasis Zuckerberg placed on AI systems of content moderation therefore has to be understood as the proactive component of a larger scheme. This effort, in turn, is intended to focus ongoing discussions about the "supervision" of digital platforms on concerns over ethics and governance—what

might be or ought to be done—while avoiding the social issues relating more directly to control, i.e., who holds the power to enforce decisions. Indeed, we saw on a few occasions Zuckerberg's obvious discomfort when interrogated about his platform's dominant position. Perhaps the best expression of this uneasiness happened when an elected official inquired: "If I buy a Ford, and it doesn't work well and I don't like it, I can buy a Chevy. If I'm upset with Facebook, what's the equivalent product I can go sign up for? . . . You don't feel like you have a monopoly?"—to which the CEO succinctly answered: "It certainly doesn't feel that way to me" (quoted in Dayen, 2018).Moreover, a leaked photo of Zuckerberg's preparation notes revealed that if pressed on the issue, he was ready to argue for the importance of strong American digital players to counterbalance the growing influence of Chinese corporations (Cadwalladr, 2018). Thus, the CEO's hesitancy to engage on this range of issues hints at the way Facebook intends to capitalize on AI solutionism's discursive tropes—i.e., how it is construed as *the* fix for any problem encountered—in order to disarm any other lines of criticism. As an answer to fake news, terrorist propaganda, and increasing Chinese economic power, "the AI tools of the next 5 to 10 years" are to be (rhetorically) scaled-up as a countermeasure against any further accusations.

As mentioned in the first section I, if legitimacy is constructed and thus performative, it relies as well on how it is received, i.e., believed and assented to. Legitimacy, simply stated, is a symbolic give and take—a *pas de deux* that needs an audience for it to develop. Inasmuch as we interpret the elected officials' poor performance during Zuckerberg's hearings as a surrendering of their responsibility with regards to the regulation of social media platforms, it is all the more interesting how Facebook's CEO seemed, in the months leading up to his testimony, to have developed a newfound acceptance toward the matter. He explained: "My position is not that there should be no regulation. . . . I think the real question . . . is what's the *right* regulation?" (quoted in Ward, 2018; emphasis added). Throughout his testimony, he expressed support for better transparency, while even vaguely recognizing the value of the EU's GDPR's principles. So what is it then? What does "right" mean, knowing that it could very well be one of the most ambiguous terms ever coined? On the one hand, it certainly resonates with the fact that Facebook and Zuckerberg are pragmatic actors here; they understand that they have no choice but to collaborate with authorities. To be sure, *realpolitik* is cybernetic. On the other, and in line with such pragmatism, it appears that consent to regulation could be seen as the best way to have a say in its elaboration. Zuckerberg's testimony

indeed allowed for the CEO of a company responsible for what has been called a crisis of democracy to be handed, *precisely because of his role in this crisis*, a leading role in the definition of the upcoming regulatory framework tasked with figuring a way out of it.

Unlawful territories; or how to cope with killer robots

The Montreal Declaration and Zuckerberg's hearings demonstrate how actors involved in AI come to deal with the many ambiguities they often bring to the fore, and how they ultimately circumvent the scant regulatory efforts laid out by governing entities. An inquiry into lethal autonomous weapon systems (LAWS) and the way they're currently debated and problematized reveals an even fuzzier picture where conflicts are both internal and external, and where unsettled problems remain exactly that: *problematic*. An increasingly controversial topic throughout the 2010s, LAWS appeared as machine-learned systems of varying degrees of autonomy allowing for the automation of context awareness, targeting, and (potentially) the lethal decision to open fire. As of April 2018, "at least 381 partly autonomous weapons and military systems" had been deployed or were in development in twelve countries (Busby, 2018); a development that entails for the field of AI a full range of reconsiderations of admitted delineations of accountability and liability. Significantly, this redefinition of the moral landscape occurred not only with regards to the question of "who took the shot" but also with respect to the responsibility one should attribute to the developers of such systems, i.e., "who allowed in the first place for such a capability to be developed?" Further complicating this last concern are two aspects of current AI development. On the one hand, machine-learned systems have "operational versatility"; for instance, computer vision tools can just as easily be used in disaster-recovery efforts as in combat missions. On the other hand, the well-known emphasis of AI scientists on applying Open Science principles (LeCun, 2018) has resulted in a drastically enhanced circulation of the concerned technologies, rendering any attempt at controlling the dissemination of LAWS all the more difficult.

Confronted with such a wide array of questions about the responsibilities incurred through what happens both on the battlefield and in the laboratory, the field of AI has been, and is still currently, undergoing a fairly significant crisis. The normative landscape making up for what should be a legitimate proximity with LAWS-related activities is changing constantly, allowing very few actors to stay outside the debate. It

is our contention that such a massive overture to ethical reinterpretations has generated a profusion of problematizing undertakings of various sizes and configurations, all of which intended to redraw the boundaries of moral and legal liability. This effervescence was ultimately induced by a general concern over not being the one actor who didn't do enough to prevent the bleak outcomes many are expecting.

As just mentioned, such translation-problematizing efforts have taken many forms that are often guided by competing interests. Among the most publicized cases of the sort were the public protests of workers from Google and Microsoft against the contracts these firms had signed with US military entities (Horgan, 2018). In both cases, employees construed the projects they were working on as going too far and crossing all sorts of lines by allowing for enhanced lethal capabilities. While Google admitted that Project Maven was perhaps straying too far from its "Don't Be Evil" mantra, it still argued that *some forms* of collaboration with the military were acceptable. For its part, Microsoft simply rebuffed its workers' claims to moral high ground by to the virtues of patriotism in a now well-known declaration: "We're not going to withhold technology from institutions that we have elected in democracies to protect the freedoms we enjoy" (Finley, 2019).

These two well-publicized cases thus cultivated a narrative of AI workers somewhat opposing collaboration with the military, but such an understanding shouldn't be generalized to the whole tech industry. As the deputy security of defense responsible for Project Maven said after the project's demise, "the department was concerned that Google might be the canary in the coal mine, but that's not what happened" (Work quoted in Simonite, 2019). In fact, collaboration between tech and the military is growing fast, and the latter appears to understand quite well how it needs to modulate its interactions with this new talent pool. When questioned about the efforts deployed by the Air Force to facilitate collaboration with tech startups, an official explained how "flexible" moral costs were a necessary component of their approach: "If you don't want to work with us on weapons systems but you do want to work on medicine or green energy or data analytics, we should have an open door that fits the needs of the partnership" (Roper quoted in Thompson, 2019). This expresses quite well the general pattern of problematizing activities and boundary-drawing currently unfolding. Many AI specialists appear intent on collaborating with military actors as long as it doesn't mean working on *explicitly* lethal systems. Thus, a figure as strongly associated with ethical AI as Bengio has said that "there's nothing to stop us from building defensive technology. There's a big

difference between defensive weapons that will kill off drones, and offensive weapons that are targeting humans" (quoted in Knight, 2018). Developing automated targeting and shooting systems is obliquely legitimated following the specious distinction between "defensive" and "offensive" systems, which inflects in an important way the topography of the still-emergent and ever-morphing normative landscape of LAWS.

While this massive enterprise of justification and boundary-setting was unfolding mostly between tech firms and AI workers, some of these very same actors were taking part in another distinct, if intersecting, field of problematizations. In effect, corporations, their mostly anonymous rank-and-file workers, and the more recognized figures of the field of AI all engaged in what could be described as a cottage industry of declarations, petitions, and open letters stating their principled rejection of LAWS. Simply through their titles, we can observe the escalating severity of these rejections as the years go by, from the Research Priorities for Robust and Beneficial Artificial Intelligence (Russell et al., 2015) to the Call for an International Ban on the Weaponization of Artificial Intelligence (Kerr et al., 2017) and, most recently, the Lethal Autonomous Weapons Pledge (Future of Life Institute, 2018). In most cases, the same names reappear, with Bengio and Stuart Russell being two main figures; but what should also be noticed is the numerous signatures from workers at groups pertaining to Google, Microsoft, and a number of other AI developers who are no stranger to military contracting. These public statements should in turn be associated—and seen as taking part in and integrating with—the more recent and well-known Campaign to Stop Killer Robots (2019). While tech workers and AI scientists were dealing with the ethical challenges of *where to draw the line not to cross*, the coalition of academic and NGO actors—Human Rights Watch, for instance—that constituted the Campaign intended to garner sufficient support from state and civil society entities to produce an internationally enforced and legally binding resolution banning the use of LAWS (Motoyama, 2018). This diplomatic pathway has for now largely failed. The initiative seems increasingly late to the game when faced with the fact that, as mentioned, at least a dozen countries are already deploying LAWS. Quite simply, the strategic benefits related to what many are calling "the third revolution in warfare" are too tantalizing for major geopolitical players to refrain from developing these systems.

So what can we expect to result from these translations when all is said and done? Many experts, from both the computer and social sciences, agree on the inevitability of some sort of LAWS eventually becoming an enduring component of military operations

(Teffer, 2018)—if they aren't already. In that regard, what remains to be determined is the degree of autonomy we'll find in such machines. Even by these standards, fully automated defensive systems, such as those supported by figures like Bengio, could easily facilitate the sort of escalation between actors like the US and China that leads to the grim outcome of fully-automated *offensive* systems. The Campaign to Stop Killer Robots and the numerous petitions, pledges, and other declarations that allowed the AI scientists and workers who signed them to position themselves as the *Résistance* against LAWS's deployment will probably only slow this outcome, while inscribing all of them in the history books as some sort of equivalent to the physicists who morally resisted the Cold War's escalation after having produced the first nuclear bomb. This is not to imply that this outcome was the intended one or that the Campaign or other initiatives were intently instrumentalized to this end. Rather, through these numerous attempts at problematizing LAWS and defining the boundaries of their legitimacy, such a dialogue ended up occupying the space that could have been used to interrogate the deeper, structural reasons why such militarized technological developments emerged in the first place. Instead of focusing on the degree of autonomy that should be granted to lethal machines, the question that could be asked relates to the socioeconomic and sociopolitical contexts that have led us to envision these innovations as inevitable, including the very effect of Open Science with regard to LAWS and how this model's embrace by the most-regarded scientists in the field never ceases to be problematic.

Conclusion

Among the very first issues raised in this piece were questions related to the different ways by which AI compelled a renewal of the social sciences. Especially significant were challenges addressed to STS scholars and how they can broaden their—critical—scope while ensuring the pursuit of more precise and empirically informed research. It is our contention that what is at stake in the foregoing is the very possibility of Critical AI Studies (CAIS). What would this entail? Specifically, would such an intellectual endeavor be up to the many challenges posed by the fast-paced and highly scalable deployment of AI technologies in increasing aspects of social life? Without a doubt, AI implementation is a total social phenomenon *à la* Mauss; whereas the wide range of its practical aspects has far-reaching consequences, the same could be said about its symbolic effects as they assemble to inform a culturally rich and mythical language of its own. For CAIS itself, this raises questions that unfold along the same path as three of the major concepts we've developed: (i) translation, (ii) criticism, and (iii) cybernetic adaptation. The discourses stemming from the social sciences and humanities are second-order discourses: they reconstruct, i.e., translate what is there in different contexts. Knowing then that CAIS is necessarily a translation (not an absolute but an interpretative one) creates the condition under which it becomes possible to aim at fair analysis and finding the *mot juste*. In turn, this allows us to better understand the critical capacity of CAIS while also clarifying how it plays out in terms of reflexivity and its ability to make sense of justificatory discourses. Following Boltanski and others, one of the arguments frequently presented above is that criticism and justification go hand in hand in a way that points toward the persistent need to propose robust yet refined analyses, which brings us to the issue of adaptation, or the need to be pertinent, both scientifically and socially. Because there are risks in either being too involved or too distanced from the debates surrounding the deployment of AI, the question of what the appropriate distance actually is remains.

From the preceding, and recognizing now that self-doubt might very well be constitutive of CAIS, some affirmative conclusions can nonetheless be offered. One deals with the theoretical quarrels of the past that were dismissed for not being helpful in and of themselves—nor reflecting the various contingent realities of AI deployment. Rather, we have tried to emphasize the continuum and fruitful discussions that exist between, for instance, the ideas of translation, spokespeople, problematization, justification, legitimacy-building, and adaptability as a type of rational as well as pragmatic power dynamic. These concepts convey much meaning as they resonate together with the social, cultural, economic, and political evolution of AI technologies—something that indeed informs this other conclusion regarding the importance of *in situ* investigations. The Montreal Declaration, Zuckerberg's hearings, and the wave of problematizations surrounding LAWS are three distinct cybernetic realms, to be sure. While their stakes, scopes, and *modi operandi* differ in many respects—interpreted as de Certeau's disparate yet commensurable and interrelated *multiple*—they allow us to more fully comprehend how the myth of AI acts as a means of social engineering, just as the prospect of AI social engineering appears as an unfolding myth. Common to these three instances of AI deployment is how they reveal vagueness and ambiguity to be highly performative, how contextualization is shown to be neither properly local nor global, and how notions like "governance" and "ethics" should be seen as instrumentally empty signifiers. Moreover, it is only together that they can test the broader

assumptions made throughout this piece and thus circle back to inform the theory put forward.

The first of these assumptions is twofold: AI translators are fundamentally unequal, and the current state of problematization is best characterized as its own solutionism. Of course, it would be difficult to compare the charisma of figures like Bengio or Zuckerberg, yet entrepreneurs. Both of them are at their best when speaking about the *promise* of a better tomorrow. In their respective problematizing ventures, they mobilize symbolic as well as financial and political resources in order to circumvent problems they barely acknowledged in the first place. To adapt, then, means to be quite vague on what is currently unfolding. The Montreal Declaration and the Facebook congressional hearings show striking similarities in this regard, whereas the more complex and unsettled dispute over LAWS remains too murky and obscure for these dynamics to transpire. Here, a fine-grained analysis allows us to engage with the hidden costs of such ambiguity. For actors who are not stars *à la* Bengio or Zuckerberg, agenda setting and the pursuit of higher principles face many obstacles—arising either from companies, the community of AI developers, or the very logic of the field's development—and thus comes to be Sisyphean tasks.

Another key assumption throughout this piece relates to the fundamental discrepancy between criticism of AI and its justification, namely, how the latter never ceases to benefit from the many weaknesses of the former. The three case studies prove deeply coherent here, notwithstanding the obvious differences of context. Dissident voices taking aim at AI deployment in Quebec are scattered and still unable to compete with the "official" discourse of star translators, and most of, if not all, the main institutions in the province. US officials tasked with pressing Facebook on its many shortcomings ended up letting Zuckerberg position his platform as a major actor in any regulatory enterprise that might emerge. Even in the debates surrounding LAWS, efforts to conceptualize risks, threats, and dangers have remained vague—but of a nonperformative and inefficient kind this time—and easily dismissed by the *realpolitik* attitude displayed by the majority of actors in the field of AI.

Lastly, we have demonstrated how the current deployment of AI technologies is indicative of an enhanced form of cybernetics where control joins communication *on behalf of* a constant need to adapt. Once again, the Montreal Declaration, Zuckerberg's testimony, and the debates surrounding LAWS are clear illustrations of the dynamic in question. While actors modulate their discourses following what they perceive to be the constraints coming from their environment, they intend just as much to adjust and manage the expectations arising from it. It is a question of both magnitude and pace. The various vectors through which AI systems are currently being deployed are transforming many practical aspects of everyday life in part because they are intertwined with more symbolic ones. It is this wider reach, in turn, which shall prompt us to accelerate the development of CAIS as an alternative and always-to-be-continued means to translate, critique, and adapt today's world.

Declaration of conflicting interests

Funding

Notes

1. Admittedly, Latour went to great lengths to repudiate most of prior social science scholarship: "Overall and in the details, social theory has failed on science *so radically* that it's safe to postulate that it had *always failed* elsewhere as well" (2005: 94).
2. Of late, this argument for a holistic, crosspollinated approach to AI and society has been made by many, such as Sudmann (2018).
3. "All actors and intermediaries have differential access to transcoding, social capital and positioning in the networks [making it] possible to prioritize the resources and constraints of action of each of them" (Lascoumes, 1996: 338; our translation).
4. The multiple ways by which the technical opacity of machine learners and broader cultural ambiguities are linked have been studied by scholars such as Burrell (2016) and Mackenzie (2015).
5. "Performers in the present innovate, create, and struggle for social change through small but significant revisions of familiar scripts which are themselves carved from deeply rooted cultural texts" (Alexander et al., 2006: 15).

ORCID iD

Marius Senneville https://orcid.org/0000-0003-1482-1136

References

Akrich M, Callon M and Latour B (2002a) The key to success in innovation part I: The art of interessement. *International Journal of Innovation Management* 6(2): 187–206.

Akrich M, Callon M and Latour B (2002b) The key to success in innovation part II: The art of choosing good spokespersons. *International Journal of Innovation Management* 6(2): 207–225.

Alexander JC, Giesen B and Mast JL (eds) (2006) *Social Performance: Symbolic Action, Cultural Pragmatics, and Ritual*. New York: Cambridge University Press.

Boltanski L (1990) Sociologie critique et sociologie de la critique. *Politix* 3(10): 124–134.

Boltanski L and Thévenot L (2006) *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Borup M, et al. (2006) The sociology of expectations in science and technology. *Technology Analysis & Strategic Management* 18(3–4): 285–298.

Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*. Available at: http://journals.sagepub.com/doi/10.1177/2053951715622512.

Busby M (2018) Killer robots: Pressure builds for ban as governments meet. *The Guardian*. Available at: www.theguardian.com/technology/2018/apr/09/killer-robots-pressure-builds-for-ban-as-governments-meet (accessed 7 April 2019).

Cadwalladr C (2018) Five questions Mark Zuckerberg should be asked by congress. *The Guardian*. Available at: www.theguardian.com/technology/2018/apr/11/mark-zuckerberg-congress-testimony-five-questions-ask-facebook-ceo (accessed 3 April 2019).

Callon M (1981) Struggles and negotiations to define what is problematic and what is not: The socio-logic of translation. In: Knorr-Cetina K, et al. (eds) *The Social Process of Scientific Investigation*. Dordrecht: Reidel, pp.197–219.

Callon M (1984) Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review* 32(1): 196–233.

Campaign to Stop Killer Robots (2019) All action and achievements. Available at: www.stopkillerrobots.org/action-and-achievements/ (accessed 1 April 2019).

Cardon D, et al. (2018) La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux* 211(5): 173–220.

Crawford K (2016) Can an algorithm be agonistic? Scenes of contest in calculated publics. *Science, Technology & Human Values* 41(1): 77–92.

Dayen D (2018) The Senate is afraid to govern. That's great news for Facebook. *The Intercept*. Available at: https://theintercept.com/2018/04/10/facebook-the-senate-is-afraid-to-govern-thats-great-news-for-facebook/ (accessed 6 April 2019).

Elish MC and boyd D (2018) Situating methods in the magic of big data and AI. *Communication Monographs* 85(1): 57–80.

Finley K (2019) Microsoft CEO defends army contract for augmented reality. *Wired*. Available at: www.wired.com/story/microsoft-ceo-defends-army-contract-augmented-reality/ (accessed 8 April 2019).

Future of Life Institute (2018) Lethal autonomous weapons pledge. Available at: futureoflife.org/lethal-autonomous-weapons-pledge/ (accessed 7 April 2019).

Gagné JF (2017) What happens when noise becomes information? *jfgagne*. Available at: jfgagne.ai/blog/2017-5-11-/ (accessed 31 March 2019).

Gillespie T (2010) The politics of 'platforms.' *New Media & Society* 12(3): 347–364.

Gillespie T (2017) Regulation of and by platforms. In: Burgess J, Poell T and Marwick A (eds) *SAGE Handbook of Social Media*. London: Sage, pp.254–278.

Greene D, et al. (2019) Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, Hawaii, USA.

Guggenheim M and Potthast J (2012) Symmetrical twins : On the relationship between Actor-Network theory and the sociology of critical capacities. *European Journal of Social Theory* 15(2): 157–178.

Guice J (1999) Designing the future: The culture of new trends in science and technology. *Research Policy* 28: 81–98.

Halpern O (2014) Cybernetic rationality. *Distinktion: Journal of Social Theory* 15(2): 223–238.

Horgan J (2018) Antiwar movement spreads among tech workers. *Scientific American*. Available at: https://blogs.scientificamerican.com/cross-check/antiwar-movement-spreads-among-tech-workers/ (accessed 8 April 2019).

IA Responsible (2018) Montreal Declaration for a responsible development of artificial intelligence. Available at: www.montrealdeclaration-responsibleai.com/ (accessed 4 March 2019).

Kerr I, Bengio Y, Hinton G, et al. (2017) Call for an International Ban on the Weaponization of artificial intelligence. *Centre for Law, Technology and Society*. Available at: https://techlaw.uottawa.ca/bankillerai#letter (accessed 5 April 2019).

Kitchin R (2014) Big data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 205395171452848.

Knight W (2018) One of the fathers of AI is worried about its future. *MIT Technology Review*. Available at: www.technologyreview.com/s/612434/one-of-the-fathers-of-ai-is-worried-about-its-future/ (accessed on 6 April 2019).

Lascoumes P (1996) Rendre gouvernable : De La 'traduction' Au 'transcodage'. L'analyse des processus de changement dans les réseaux d'action publique. In: Chevallier J (eds) *La Gouvernabilité*, Paris: PUF, pp.325–338.

Lash S (2007) Power after hegemony: Cultural studies in mutation? *Theory, Culture & Society* 24(3): 55–78.

Latour B (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

LeCun Y (2018) Facebook's chief AI scientist says that silicon valley needs to work more closely with academia to build the future of artificial intelligence. *Business Insider*. Available at: www.businessinsider.com/facebook-yann-lecun-dual-affiliation-model-ai-experts-2018-8 (accessed 7 April 2019).

Levin S (2019) 'Bias deep inside the code': The problem with AI 'ethics' in Silicon Valley. *The Guardian*. Available at: www.theguardian.com/technology/2019/mar/28/big-tech-ai-ethics-boards-prejudice (accessed 27 March 2019).

L'Heureux A, Grolinger K, ElYamany HF, et al. (2017) Machine learning with big data: Challenges and approaches. *IEEE Access* 5: 7776–7797.

Mackenzie A (2015) The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.

Mackenzie A (2017) *Machine Learners: Archaeology of a Data Practice*. Cambridge, MA: MIT Press.

Merton RK (1948) The self-fulfilling prophecy. *The Antioch Review* 8(2): 193–210.

Metz C (2019) Turing award won by 3 pioneers in artificial intelligence. *The New York Times*. Available at: www.nytimes.com/2019/03/27/technology/turing-award-ai.html (accessed 2 April 2019).

Morozov E (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York: Public Affairs.

Motoyama S (2018) Inside the United Nations' effort to regulate autonomous killer robots. *The Verge*. Available at: www.theverge.com/2018/8/27/17786080/united-nations-un-autonomous-killer-robots-regulation-conference (accessed 14 March 2019).

Natale S and Ballatore A (2017) Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of Research into New Media Technologies*. Available at: http://journals.sagepub.com/doi/10.1177/1354856517715164

Newton C (2018) The 5 biggest takeaways from Mark Zuckerberg's appearance before the senate. *The Verge*. Available at: www.theverge.com/2018/4/10/17222444/mark-zuckerberg-senate-hearing-highlights-cambridge-analytica (accessed 15 March 2019).

Rettino-Parazelli K (2017) L'intelligence artificielle, moteur économique. *Le Devoir*. Available at: https://www.ledevoir.com/economie/500340/dominique-anglade-au-devoir (accessed 5 April 2019).

Ricœur P (1998) *Essais D'herméneutique. 2: Du Texte à L'action*. Paris: Seuil.

Roberge J (2011) What is critical hermeneutics? *Thesis Eleven* 106(1): 5–22.

Roberge J (2016) Hermeneutics and cultural sociology. In: Inglis D (ed) *The Sage Handbook of Cultural Sociology*. London: Sage, pp.272–281.

Roberge J, Morin K and Senneville M (2019) Deep learning's governmentality: The other black box. In: Sudmann A (eds) *The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms*. Bielefeld: Transcript, pp.123–142.

Roberge J and Seyfert R (2018) What are algorithmic cultures? In: Seyfert R and Roberge J (eds) *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies*. New York: Routledge, pp.1–25.

Rosenberg M, et al. (2018) How Trump consultants exploited the Facebook data of millions. *The New York Times*. Available at: www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html (accessed 28 March 2019).

Russell S, Dewey D and Tegmark M (2015) Research priorities for robust and beneficial artificial intelligence. *AI Magazine* 36(4): 105–114.

Simonite T (2019) The Pentagon doubles down on AI – and wants help from big tech. *Wired*. Available at: www.wired.com/story/pentagon-doubles-down-ai-wants-help-big-tech/ (accessed 3 April 2019).

Sudmann A (2018) On the media-political dimension of artificial intelligence: Deep learning as a black box and OpenAI. *Digital Culture & Society* 4(1): 181–200.

Teffer P (2018) AI experts tell MEPs ban on 'killer robots' is unrealistic. Available at: https://euobserver.com/science/143086 (accessed 26 March 2019).

Thompson N (2019) The air force wants to give you its credit card. *Wired*. Available at: www.wired.com/story/will-roper-air-force-interview/ (accessed 4 April 2019).

Wagner B (2018) Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In: Hildebrandt M (eds) *Being Profiling. Cogitas Ergo Sum*, Amsterdam: Amsterdam University Press, pp.1–7.

Ward A (2018) Watch Sen. Graham Grill Zuckerberg on whether Facebook is a monopoly. *Vox*. Available at: www.vox.com/2018/4/10/17220878/facebook-mark-zuckerberg-testimony-monopoly-graham (accessed 18 March 2019).

Zuckerberg M (2017) Building global community. Available at: www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/ (accessed 17 March 2019).

Zuckerberg M (2018a) Transcript of Mark Zuckerberg's senate hearing. *The Washington Post*. Available at: www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?utm_term = .671ba216aa31 (accessed 16 March 2019).

Zuckerberg M (2018b) Transcript of Zuckerberg's appearance before house committee. *The Washington Post*. Available at: www.washingtonpost.com/news/the-switch/wp/2018/04/11/transcript-of-zuckerbergs-appearance-before-house-committee/?utm_term = .939c0d20484b (accessed 16 March 2019).